

论文题目：人工智能算法决策与企业研发“合谋”

作者：

徐雷，经济学博士，辽宁大学经济学部经济学院教授。主要研究领域：产业组织与博弈论、算法合谋与反垄断。联系电话：13841679860（微信同步）；E-mail: [xulei@lnu.edu.cn](mailto:xulei@lnu.edu.cn).

# 人工智能算法决策与企业研发“合谋”

徐雷

(辽宁大学经济学部经济学院, 辽宁沈阳, 110036)

**摘要:** 文章设定了一个企业研发博弈的智能体决策环境, 运用强化学习 Q-learning 算法对多智能体进行联合训练, 分析和预测由人工智能决定企业研发投入的可能结果。主要结论如下: 第一, 尽管智能体决策实现“合谋”结果的概率并不高, 但与纳什均衡结果相比, 研发投入更低而利润水平更高的情况是普遍存在的。第二, 智能体数量的增加能够降低“合谋”的实现概率, 且这种影响具有较高的全局性。第三, 研发的外溢程度会促进智能体“合谋”, 但这种促进作用是局部的, 在智能体数量较多时更易实现。第四, 学习率的提高能够降低“合谋”的实现概率, 但这种作用也仅在局部发生, 在智能体数量较多且研发外溢性较低时作用更显著。基于以上研究结果, 文章提出了加强人工智能算法在企业决策层面应用管理的政策建议, 为数字经济时代反垄断规制的完善提供了理论支持。

**关键词:** 人工智能; Q-learning 算法; 研发博弈; 企业创新

**中图分类号:** F406.3

## 一、引言

在新科技革命浪潮下, 企业作为重要的创新主体, 加大研发投入已成为其增强核心竞争力的重要手段, 也是推动我国经济实现高质量发展的重要路径。但是, 对企业来讲, 研发活动是一柄双刃剑, 一方面, 研发成功能够有效提升企业竞争力; 但另一方面, 一旦研发失败, 则可能导致产品更新不及时、资金周转困难, 甚至导致企业运营危机。因此, 为了规避风险、提升研发活动的规模经济性, 企业往往寻求研发合作, 这种研发合作能够促进创新、提高效率、扩大社会福利, 但也会导致企业合谋从而阻碍科技进步<sup>[1]</sup>。Sovinsky<sup>[2]</sup>指出了研发合资公司 (Research Joint Ventures) 导致成员企业共谋的几种方式, 包括促进集中决策, 限制竞争条款, 促进信息交流等。尽管学界对于企业共同研发的规制和豁免政策存在一定争议<sup>[3,4]</sup>, 但企业共同研发行为的辨识性是明确的, 能够受到反垄断法明确的规制和约束。

然而, 随着数字经济时代的到来, 企业在重要决策中越来越多地使用了计算机程序辅助。其中的一些程序利用了人工智能机器学习 (Machine Learning) 算法, 能够在自己过去的行动经验中自主学习, 这引起了我们对这种人工智能决策是否会在没有事前沟通的条件下形成“合谋”的关注, 而这将对现有的反垄断法形成重要挑战。以我国反垄断法为例, 《中华人民共和国反垄断法》<sup>①</sup>第一章第三条规定“本法规定的垄断行为包括: 一、经营者达成垄断协议; 二、经营者滥用市场支配地位; 三、具有或者可能具有排除、限制竞争效果的经营者集中。”第二章第十三条规定, 禁止具有竞争关系的经营者达成“限制购买新技术、新设备或者限制开发新技术、新产品”的垄断协议, 并明确指出“本法所称垄断协议, 是指排除、限制竞争的协议、决定或者其他协同行为。”显然, 研发合谋是垄断的一种形式, 但由机器学习算法实施的“合谋”行为具有以下两个突出的特点: 一是人工智能行为主体模糊, 不具有协同行为的主观意愿; 二是具有竞争关系的经营者 (算法) 之间没有协议甚至没有任何沟通从而使“合谋”存在极强的隐蔽性<sup>[5-7]</sup>, 导致默契合谋 (tacit collusion) 的形成<sup>[8-10]</sup>。对默契合谋是否应成为反垄断执法对象的讨论甚至早于算法合谋的出现, 一些学者认为只要不存

<sup>①</sup> 2007年8月30日第十届全国人民代表大会常务委员会第二十九次会议通过。

在协议，就不能将事实上的协同结果判定为合谋<sup>[11,12]</sup>，但更多的学者持相反的观点，他们认为反垄断法应将默契合谋界定为垄断协议，从而对其进行规制<sup>[13-17]</sup>。尽管如此，反垄断法对此种默契合谋进行取证和裁量时，仍面临责任定位困难、框架体系失衡、算法控制与算法合谋协议概念局限等诸多难题<sup>[18-21]</sup>。

为了对算法“合谋”进行更好的规制，我们首先要考察算法决策是否会导致企业的研发“合谋”，并分析算法行为选择的影响因素。当前，有两条技术路线可以验证算法“合谋”是否存在：第一条技术路线是从真实的市场结果进行判断。例如，Milgrom 和 Tadelis<sup>[22]</sup>分析了几个人工智能算法辅助决策者提升市场运营效率的案例，Assad 等<sup>[23]</sup>分析了德国的汽油零售市场，发现人工智能定价算法的采用与加油站的利润率大幅增加相关。第二条技术路线是考察人工智能算法决策的仿真结果来分析合谋的可能性。目前，这种技术路线受到了越来越广泛的关注，重要的研究成果不断出现。例如，Salcedo<sup>[24]</sup>设计了一个双寡头 Bertrand 动态竞争模型，发现定价算法会实现接近垄断价格的均衡。Zhou 等<sup>[25]</sup>研究了双寡头古诺模型中的算法合谋问题，Calvano 等<sup>[26]</sup>指出了价格竞争中智能体会实施惩罚策略，Klein<sup>[27]</sup>研究了续费价格竞争，他们的研究均表明算法决策可能导致智能体实现一定程度的合谋。Meylahn 和 Janseen<sup>[28]</sup>研究了迭代囚徒困境、猎鹿和鹰鸽博弈，在所有三种博弈中，唯一可能的均衡策略组合是“赢驻输变”（win-stay, lose-shift）策略。

那么，为何算法决策会导致默契合谋呢？这在学理主要有以下两点解释<sup>[5]</sup>：一是算法作为工具有助于均衡结果达成。Beneke 和 Mackenrodt<sup>[29]</sup>认为算法具有对不确定变量做出准确预测从而协调“超价格”定价的能力。此外，算法促进达成协调结果的能力使寡头协调价格的行为不仅发生于双寡头市场，多寡头企业也能在算法协助下协调行为<sup>[30,31]</sup>。二是算法使合谋具有可维持性。使用平行算法的经营者利用了同一规则对市场变动和对手偏离做出同一反应，能使合谋更易形成和维持<sup>[32]</sup>。由于特定贴现率波动对卡特尔维持时间有显著影响<sup>[33]</sup>，Mehra<sup>[34]</sup>指出当算法使偏离的当期收益小于合谋未来收益的贴现值之和时，合谋就具有可维持性。面对算法合谋的易发性，学者们在法学和经济学的框架下，从企业使用算法的角度<sup>[35-38]</sup>和消费者使用算法的角度<sup>[39-41]</sup>，分析了算法合谋的影响和规制问题。

本文在已有研究的基础上，考察算法决策能否导致企业研发“合谋”，我们的研究动机如下：第一，现有研究集中于在价格竞争或产量竞争中考察算法决策导致的后果，很少关注到企业研发竞争中的算法决策问题。但对我国企业来说，人工智能技术的运用受到大力支持，但它是否会促进企业研发还未可知，我们急需知道是否存在“人工智能赋能下的技术进步阻碍效应”。第二，与制定价格和制定产量的活动不同，企业的研发投入存在外溢性，这种外溢性对算法决策将产生怎样的影响，这种影响是否与博弈理论的逻辑分析相一致都是需要进行考察的重要问题。第三，从对强化学习智能体的训练上看，尽管在一个固定环境下的单智能体行为已被证明是最优化的<sup>[42]</sup>，但多智能体联合训练的行为逻辑仍然难以给出明确的理论解释<sup>[43]</sup>。因此，本文理论分析仅限于构建博弈环境以及对纳什均衡与合谋结果的比较，研究结论均是基于计算仿真结果的。但同时，我们也将对算法决策的各影响因素所发挥的作用及彼此间的协同效应进行分析，以期能够对算法决策的内在逻辑进行尝试性揭示。最后，尽管算法决策已经成为研究热点，标志性研究成果不断涌现，但中文文献却似乎忽略了这一重要方向。我们希望通过本文吸引更多的中文写作学者关注该领域，产出优秀的中文研究成果。

## 二、企业研发博弈模型设定

这里，我们构建一个研发投入具有溢出效应的企业研发博弈模型，用以分析和比较企业间在竞争和合作两种状态下的研发投入，为算法决策提供一个训练环境。

假设企业在一个同质产品市场中进行数量竞争，市场中的需求函数为：

$$P = a - Q \quad \left( Q = \sum_{i=1}^n q_i \right) \quad (1)$$

企业的研发活动是工艺创新型的，即通过研发活动，探寻降低生产特定产品成本的技术。我们以 $x_i$ 表示企业 $i$ 承担的研发表量 ( $i = 1, 2, \dots, n$ )，并以 $c_i(x_i, x_{-i})$ 表示企业 $i$ 的单位生产成本，这里的 $x_{-i}$ 为表示除 $i$ 外其他企业研发投入向量。该成本函数被设定为  $n$  个企业研发投入水平的函数，也即企业的单位生产成本不仅取决于自身的研发投入，也受到其他企业研发投入的影响。正式地，令：

$$c_i(x_i, x_{-i}) \equiv \alpha - x_i - \beta \sum_{j \neq i} x_j \quad i = 1, 2, \dots, n; j \neq i; \beta > 0 \quad (2)$$

我们假设式 (2) 中  $\beta > 0$ ，这表示企业的生产成本会随着其他企业研发投入的增加而下降，这就刻画了企业研发活动所形成的正的外溢效应。另外，研发对企业来说是有成本的，我们将企业 $i$ 投入的研究水平 $x_i$ 的成本表示为 $TC_i(x_i)$ ，并有：

$$TC_i(x_i) = x_i^2 / \gamma \quad (3)$$

下面，我们首先计算企业不合作时，它们利润最大化的研发水平。之后，再计算企业可以进行合作时，联合利润最大化的研发水平。

### 1. 非研发合作

在该博弈中，企业首先选择研发投入水平，从而确定成本函数，之后再选择产量水平。为求解博弈均衡，我们首先求解企业在产量竞争中的均衡解，然后采用逆向归纳法，求解企业的研发投入水平。

企业的产量竞争发生在研发投入之后，也即可将企业的边际成本视为既定，此时， $n$  个企业竞争的均衡产量和利润水平即为古诺均衡，由下式给出：

$$q_i = \frac{a - nc_i + \sum_{j \neq i} c_j}{n + 1} \quad (4)$$

$$\pi_i = \frac{(a - nc_i + \sum_{j \neq i} c_j)^2}{(n + 1)^2} \quad (5)$$

基于此，我们再分析企业的研发投入：每个企业在其他企业研发水平给定的情况下，以不合作的态度选定其研发水平，此时，我们要求解一个研发水平的纳什均衡。为此，将式 (2) 代入式 (5)，再减去企业的研发成本，可以得到企业的利润函数为：

$$\begin{aligned} \pi_i &= \frac{1}{(n + 1)^2} \left[ a - n \left( \alpha - x_i - \beta \sum_{j \neq i} x_j \right) + \sum_{j \neq i} \left( \alpha - x_j - \beta \sum_{-j} x_{-j} \right) \right]^2 - \frac{(x_i)^2}{\gamma} \\ &= \frac{1}{(n + 1)^2} \left[ \alpha - a - (n - (n - 1)\beta)x_i + (1 - 2\beta) \sum_{j \neq i} x_j \right]^2 - \frac{(x_i)^2}{\gamma} \end{aligned} \quad (6)$$

对一个给定的研发水平 $x_j$ ，企业 $i$ 选择 $x_i$ 以最大化其利润 $\pi_i$ ，一阶条件为：

$$0 = \frac{\partial \pi_i}{\partial x_i} = \frac{-2(n - (n - 1)\beta)}{(n + 1)^2} \left[ \alpha - a - (n - (n - 1)\beta)x_i + (1 - 2\beta) \sum_{j \neq i} x_j \right] - \frac{2x_i}{\gamma} \quad (7)$$

企业利润最大化需要满足二阶条件  $\partial^2 \pi / \partial x_i^2 < 0$ ，也即

$$\frac{n - (n + 1)/\sqrt{\gamma}}{n - 1} < \beta < \frac{n + (n + 1)/\sqrt{\gamma}}{n - 1} \quad (8)$$

由于利润函数在各个企业之间是对称的，我们可以求一个对称的纳什均衡，在该均衡下，有  $x_1 = x_2 = \dots = x_n \equiv x^{nc}$ ，这里的  $x^{nc}$  即是在非合作条件下，每个企业均衡的研发投入水平，它由下式给出：

$$x^{nc} = \frac{(a - \alpha)(2\beta + 2n - 2\beta n)}{\left( \frac{2}{\gamma} - \frac{(2\beta + 2n - 2\beta n)(\beta n - \beta + 1)}{(n + 1)^2} \right) (n + 1)^2} \quad (9)$$

## 2. 合作研发

在合作研发的情况下，企业在将产量竞争中各企业产量选择作为既定的条件下，共同选择能够使联合利润最大化的研发投入水平。即企业联合选择各自的研发投入水平  $x_i$ ，以最大化联合利润：

$$\max_{x_i} \sum_{i=1}^n \pi_i \quad (10)$$

其中  $\pi_i$  由式 (6) 给出， $i = 1, 2, \dots, n$ 。一阶条件为：

$$0 = \frac{\partial \sum_{i=1}^n \pi_i}{\partial x_i} = \frac{\partial \pi_1}{\partial x_i} + \frac{\partial \pi_2}{\partial x_i} + \dots + \frac{\partial \pi_i}{\partial x_i} + \dots + \frac{\partial \pi_n}{\partial x_i} \quad (11)$$

式中  $\partial \pi_i / \partial x_i$  项衡量企业  $i$  从研发投入  $x_i$  的微小改进获取的边际利润，其他项衡量企业  $j$  ( $\neq i$ ) 的利润的边际增加，这源于企业  $i$  的研发努力增加带来的溢出效应。因此，上式可写为：

$$0 = \frac{-2(n - (n - 1)\beta)}{(n + 1)^2} \left[ \alpha - a - (n - (n - 1)\beta)x_i + (1 - 2\beta) \sum_{j \neq i} x_j \right] - \frac{2x_i}{\gamma} + \frac{2(1 - 2\beta)}{(n + 1)^2} \sum_{j \neq i} \left[ \alpha - a - (n - (n - 1)\beta)x_j + (1 - 2\beta) \sum_{k \neq j} x_k \right] \quad (12)$$

由于利润函数在各个企业之间是对称的，我们可以求解一个对称的合作均衡，在该均衡中， $x_1 = x_2 = \dots = x_n \equiv x^c$ ，其中， $x^c$  是每个企业在合作均衡中的研发投入水平：

$$x^c = \frac{\gamma(a - \alpha - a\beta + \alpha\beta + a\beta n - \alpha\beta n)}{2\gamma\beta^2 n - \gamma\beta^2 n^2 - \gamma\beta^2 - 2\gamma\beta n + 2\gamma\beta + n^2 + 2n - \gamma + 1} \quad (13)$$

## 3. 对模型参数的数值设定

现在，我们对以上模型中的参数进行具体设定，一方面能够对非合作和合作状态下的企业行为进行更加直观的比较，另一方面为训练和分析人工智能决策提供一个环境。令  $a = 100$ ，

$\alpha = 50$ ,  $\gamma = 2$ , 依据 (8) 式给出的二阶条件, 我们设定厂商数量  $n$  分别取 2, 3 和 4, 研发外溢度  $\beta$  分别取 0.2, 0.3, 0.4 和 0.5。在这样的参数设定下, 我们计算了不同情况下厂商在非合作 (NE) 与合作 (CE) 均衡状态中的研发投入水平 (R&D)、产量水平 (Q) 和利润水平 ( $\pi$ ), 表-1 展示了计算结果。

表-1 对非合作研发与研发状态下企业研发、产量和利润水平的比较

企业数量	观测指标	$\beta=0.2$		$\beta=0.3$		$\beta=0.4$		$\beta=0.5$	
		NE	CE	NE	CE	NE	CE	NE	CE
2	R&D	38.46	19.61	37.12	23.13	35.40	27.56	33.33	33.33
	Q	32.05	24.51	32.75	26.69	33.19	29.53	33.33	33.33
	$\pi$	287.64	408.50	383.76	444.84	474.78	492.13	555.56	555.56
3	R&D	29.82	11.59	28.85	14.71	27.23	18.91	25.00	25.00
	Q	22.94	16.56	24.04	18.38	24.75	21.01	25	25
	$\pi$	81.54	206.95	161.80	229.78	242.01	262.61	312.5	312.5
4	R&D	24.08	8.05	23.45	10.69	22.08	14.36	20.00	20.00
	Q	17.71	12.58	18.91	14.06	19.72	16.32	20.00	20.00
	$\pi$	23.57	125.75	82.68	140.61	144.92	163.19	200.00	20.00

由表-1 可以看出, 在合作状态下, 企业的研发投入和产量水平比非合作状态下更小, 利润则更高。但随着研发外溢程度 ( $\beta$ ) 的提高, 二者的差距不断减小, 在  $\beta=0.5$  时, 合作与非合作状态均衡完全相同。实际上, 如果令  $\beta$  继续变大就会发现, 在  $\beta>0.5$  时, 合作状态下的研发投入和产量水平将会更高, 但利润水平仍是合作状态下更高。

那么, 如果企业间重复进行这一研发博弈, 会出现什么结果呢? 实际上, 现代博弈理论已经证明, 对于一个仅有唯一纯策略纳什均衡的博弈, 它的有限次重复博弈的纳什均衡不会发生改变。对于我们所设定的企业研发博弈, 纯策略纳什均衡仅有一个, 那么, 机器学习算法博弈的结果是否也会向纳什均衡收敛, 研发活动的外溢度对博弈结果是否存在影响, 这是我们需要回答的问题。下面, 我们对本文所选用的机器学习算法进行简要介绍, 之后展示并分析训练结果以回答这些问题。

### 三、强化学习与 Q-Learning 算法

#### 1. 强化学习与算法选择

强化学习是一种目标导向的计算方式, 在这种计算方式下, 计算机通过与未知的动态环境进行互动来学习如何完成一项任务, 即如何根据情况采取行动, 从而实现数值奖励信号最大化<sup>②</sup>。计算机完成任务的过程不会接到人类施加的动作指令, 也不需要显著式编程 (explicitly programmed), 而是通过一系列自行尝试和自主决策去发现回报最高的行动方案。图-1 给出了一个一般的强化学习框架, 它可以被概括如下: (1) 强化学习的目标是训练智能体 (agent) 在未知的环境中完成任务。(2) 智能体接受来自环境的观测状态和奖励, 并向环境输出行动。奖励是对智能体选择的行动对完成整体任务所发挥作用的一种测度。(3) 智能体包含两个组成部分, 即政策 (policy) 和学习算法 (learning algorithm)。政策是从对环境的观测状态到智能体行动选择的一个映射。通常的, 政策可被看做是一个具有可调参数的函数逼近器 (function approximator), 而学习算法的作用就是依据行动、观测状态和奖励在不断的试错过程中对政策参数进行更新, 学习算法的目标是找到在完成任务过程中能够最大化累积奖励的最优政策。

<sup>②</sup> OECD (2017) 将机器学习分为监督学习 (supervised learning)、无监督学习 (unsupervised learning) 和强化学习 (reinforcement learning) 三种类型。

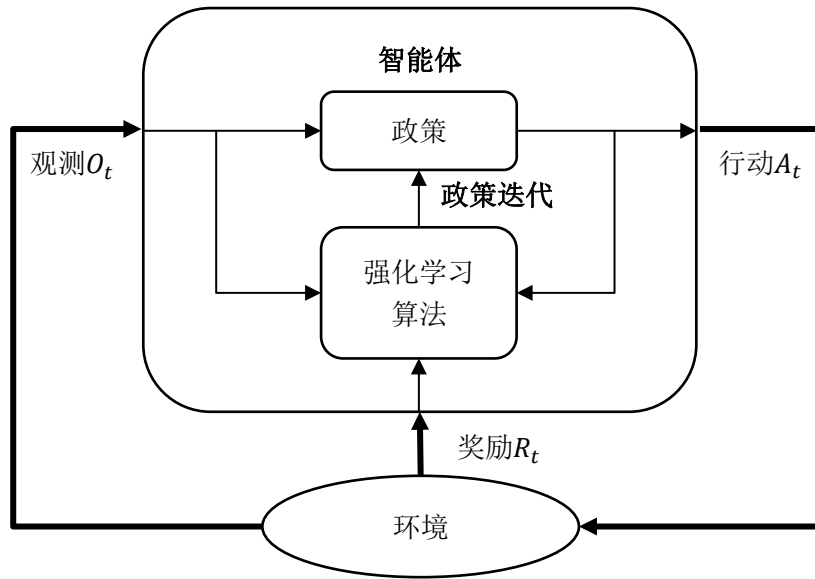


图-1 强化学习运算框架

(图片来源: [What Is Reinforcement Learning? - MATLAB & Simulink \(mathworks.com\)](https://www.mathworks.com/help/matlab/reinforcement-learning-what-is-reinforcement-learning.html))

尽管在强化学习领域有诸多学习算法可供选择,但本文基于以下几点原因选用 Q-learning 算法对智能体进行训练。第一,我们希望所使用的算法是在实践中被广泛运用的,然而,尽管目前人工智能参与企业经营决策愈加普遍,但企业却极少公布他们所使用的算法,这为我们选择合适的算法增添了难度。但是,在计算科学领域, Q-learning 算法已经得到广泛应用,鉴于该领域理论与实践具有紧密联系,我们可以相信 Q-learning 算法也会被运用到企业正在使用或即将使用的决策软件中<sup>[26]</sup>。第二, Q-learning 算法的编写和运行比较简单,运用 matlab 或 python 等软件可以通过几个参数的设定较为简易地完成代码编写,并且这些参数在经济学和博弈论的范畴内能够给予明确的解释,这意味着 Q-learning 算法更适合产业组织领域的分析使用<sup>[44,45]</sup>。第三, Q-learning 算法与更复杂的算法共享相同的架构,这些算法在很多领域取得了重要的成功,如在电子游戏对战和战棋类游戏对弈等任务中,这些算法的表现均超越了人类<sup>[46-48]</sup>,这使我们相信该算法在企业决策领域也具有很大的应用潜力。当然, Q-learning 算法也存在其缺点,比较明显的是其运算较慢,这是由其独特的迭代计算方式所导致的。

## 2. Q-learning 算法

与所有的强化学习算法相同, Q-learning 算法也是基于已有经验选择行动,也即是选择已被证明更成功的行动的概率更高,选择已被证明更不成功的行动的概率更低。通过这样的行动选择方式,智能体能够得到一个最优的政策(或趋近于最优的政策)。最初, Q-learning 算法是被用来处理马尔科夫决策过程的<sup>[49]</sup>。在一个静态的马尔科夫决策过程中,智能体在时期  $t (=1,2,\dots)$  观测到一个状态  $s_t \in S$  并选择一个行动  $a_t \in A(s_t)$ 。对于任意的  $s_t$  和  $a_t$ ,智能体会得到环境奖励  $\pi_t$ ,同时环境进入到下一个状态  $s_{t+1}$ 。 $\pi_t$  和  $s_{t+1}$  依据不随时间发生改变的条件概率分布函数  $F(\pi_t, s_{t+1} | s_t, a_t)$  产生。智能体的决策目标是最大化各期奖励的现值之和,即:

$$E \left[ \sum_{t=0}^{\infty} \delta^t \pi_t \right] \quad (14)$$

上式中,  $\delta < 1$  为折现因子。这个动态规划问题通常以贝尔曼值函数 (Bellman's value function) 的方式加以解决,即:

$$V(s) = \max_{a \in A} \{E[\pi|s, a] + \delta E[V(s')|s, a]\} \quad (15)$$

上式中,  $s'$  为  $s_{t+1}$  的简写。下面, 我们定义 Q 函数, 其具体形式为:

$$Q(s, a) = E[\pi|s, a] + \delta E[\max_{a' \in A} Q(s', a')|s, a] \quad (16)$$

不难看出, Q 函数与值函数的关系为  $V(s) \equiv \max_{a \in A} Q(s, a)$ 。因为 S 和 A 都是有限的, Q 函数的实际呈现形式为一个  $|S| \times |A|$  矩阵。

(1) 学习。如果智能体知道 Q 矩阵, 它就可以在任意给定的状态下计算出最优行动。Q-learning 的本质即是智能体在不知道概率分布函数  $F(\pi, s'|s, a)$  的情况下, 估算 Q 矩阵的一种方法。

Q-learning 算法运用迭代程序对 Q 矩阵进行估算。从一个任意的初始矩阵  $Q_0$  开始, 在状态  $t$  下选择了行动  $a_t$  后, 算法观察到  $\pi_t$  和  $s_{t+1}$  并对  $Q_t(s, a)$  矩阵中对应的元素进行更新。正式地, 对于  $s = s_t$ ,  $a = a_t$ , 算法更新 Q 矩阵对应元素的方法为:

$$Q_{t+1}(s, a) = (1 - \theta)Q_t(s, a) + \theta \left[ \pi_t + \delta \max_{a' \in A} Q_t(s', a') \right] \quad (17)$$

上式表明, 对于所要更新的矩阵元素, 其更新值  $Q_{t+1}(s, a)$  是上期值和当期奖励与当期行动后所要到达的 Q 矩阵元素的贴现值  $Q_t(s', a)$  的凸组合。对于所有其他矩阵元素  $s \neq s_t$ ,  $a \neq a_t$ , Q 值不发生改变, 即有  $Q_{t+1}(s, a) = Q_t(s, a)$ 。式中的权重  $\theta \in [0, 1]$  被称为学习率 (learning rate)。

从 Q-learning 算法估算 Q 矩阵的方式可以看出, 智能体在某种意义上是存在记忆的, 但这种记忆与博弈论中所定义的“完美记忆”存在明显差异。所谓的完美记忆指的是博弈中的参与人知道博弈任何阶段中所发生过的事实。也只有在完美记忆的假设下, 博弈参与人才能为重复博弈制定完备的策略集, 从而依据对纳什均衡的判断在策略集中选择策略并依据所选策略采取行动。但在 Q-learning 算法中, 智能体不记得之前的每一次博弈自己和其他智能体是怎样行动的, 以及获得了怎样的收益, 因此, 智能体不具有制定指导行动选择的策略的能力, 也不能在重复博弈中对不合作的参与人采取处罚行动。之前博弈的记忆全部凝结在 Q 矩阵中, 即之前的每次行动都对 Q 矩阵进行了更新, 而当前的行动选择是依据当前的 Q 矩阵进行的。因此, 智能体的行动选择逻辑与博弈论中参与人的行动选择逻辑是完全不同的。实际上, 现有文献已经证明, 当采用单智能体训练 (independent learning) 方式时 (即将其他参与人的行动选择作为环境的一部分), 智能体的行动选择会收敛到纳什均衡, 但单智能体训练显然与实际应用场景不符, 因此本文运用多智能体联合训练 (joint learning) 的方式。

(2) 试验。为了从一个任意的初始矩阵  $Q_0$  向真实的 Q 矩阵逼近, 需要在所有状态下尝试所有的行动。因此, 这种学习过程是非常耗时的, 这就需要在继续学习和利用已经获得的知识之间进行权衡<sup>③</sup>。我们希望得到的理想状态是找到这种权衡的最优解, 但这却十分困难, Q-learning 的解决方式是外生性地给定权衡方式。

本文采用的权衡方式是得到广泛使用的  $\epsilon$ -贪婪 ( $\epsilon$ -greedy) 探索模型<sup>④</sup>。依据该模型,

<sup>③</sup> 由于 Q-learning 每次计算智能更新一个矩阵要素, 而收敛到真实的 Q 矩阵要求每个要素需要被更新多次, 这导致学习过程非常耗时, 并且状态和行动空间越大, 所需的迭代次数和计算时间就会越高。

<sup>④</sup> 实际上, 在 Q-learning 算法中还有更加复杂的探索模式的计算方法, 例如, Boltzman 探索模型 (参见 Kaelbling 等, 1996<sup>[50]</sup>; Sandholm 和 Crites, 1996<sup>[51]</sup>) 也在 Q-learning 智能体训练中得到较为广泛的应用 (如 Waltman 和 Kaymak, 2008<sup>[43]</sup>), 但本文根据所研究的内容选择  $\epsilon$ -贪婪模型, 能够使智能体的训练过程更加简洁和便于理解。



学习算法以 $\varepsilon$ 的概率在所有行动中进行均匀化的随机选择，以固定的概率  $1-\varepsilon$ 选择当前的最优行动（即在对应状态下具有最高 Q 值的行动），即

$$a = \underset{a}{\operatorname{argmax}} Q(s, a) \quad (18)$$

因此，在学习过程中， $1-\varepsilon$ 就是算法采用优选模式（exploitation mode）的概率，而 $\varepsilon$ 是算法采用探索模式（exploration mode）的概率。随着学习次数的增加，智能体对环境和收益的了解不断增强，采用探索模式的概率应该不断下降，因此，我们设定一个 $\varepsilon$ 的消减率为 $\xi$ ， $\varepsilon$ 随时间的变化由该消减率定义，即有：

$$\varepsilon(t) = \varepsilon \cdot (1 - \xi)^t \quad (19)$$

上式表明，在初始期，智能体采用探索模式的概率为 $\varepsilon$ ，随着学习的不断进行，这一概率会不断减小，直至趋近于 0。但在实际的智能体训练过程中，往往还要为 $\varepsilon$ 设定一个最低的 $\underline{\varepsilon}$ 值，我们将其记为 $\underline{\varepsilon}$ ，随着训练过程的进展使得 $\varepsilon$ 低于 $\underline{\varepsilon}$ 时，智能体将以 $\underline{\varepsilon}$ 作为采用探索模式的概率，也即保证智能体始终存在探索的机会。

#### 四、智能体训练结果

本文考察人工智能在研发博弈中的行为选择，与现有文献对价格和产量博弈的诸多探讨不同的是，企业的研发投入存在外溢性，因此，本文首先基于外溢度 $\beta$ 的不同设定对智能体进行分别训练。依据前文， $\beta$ 在大于 0.5 时企业合作均衡下的研发投入水平和产量水平大于非合作均衡，这与现实状态不符，因此，我们仅考虑 $\beta$ 小于 0.5 的情况。另外，若 $\beta$ 小于 0.2，当企业数量大于等于 4 时企业利润最大化的二阶条件（8）式无法得到满足，因此，我们排除 $\beta$ 小于 0.2 的情况。最终，我们将 $\beta$ 分别设定为 0.2，0.3 和 0.4 三种情况，同时，企业数量分别设定为 2，3 和 4 家。另外，借鉴现有文献，我们令 $\delta = 0.9$ ， $\varepsilon = 0.9$ ， $\underline{\varepsilon} = 0.01$ ，学习率 $\theta$ 则分别设定为 0.25，0.50 和 0.75 三种不同情况。

依据 $\beta$ ， $n$ 和 $\theta$ 的不同组合，我们设定了 27 种不同的智能体训练环境，由于随机数对智能体结果产生影响，因此我们在每种训练环境下对智能体进行 100 次训练。为了在每次训练中能够让智能体有足够的机会尝试每一种策略选择和策略组合，我们将每次训练的片段数（episode）设定为策略组合数量的 100 倍。在每次训练中，如果最后一个片段的行动结果与最后 50 个片段的平均行动结果相差小于 1%，则认为训练结果收敛而可以接受，否则不接受此次训练结果，最终的行动结果即为最后 50 个片段的平均结果。

为了更直观地考察智能体的行为选择更接近于非合作均衡还是合作均衡，我们设定了“合谋指数（CI）”，其形式为：

$$CI = \frac{V_{AI} - V_{NC}}{V_C - V_{NC}} \quad (20)$$

具体的，我们设定了两个合谋指数，即研发投入合谋指数（ $CI_{rd}$ ）和利润水平合谋指数（ $CI_{pf}$ ）。始终， $V_{AI}$ 表示所考察标量（研发投入或利润水平）的智能体训练结果值， $V_{NC}$ 表示纳什均衡值， $V_C$ 表示合谋结果值。因此，该指数取值在[0,1]区间内，越接近 0 意味着人工智能决策越接近于纳什均衡，越接近于 1 意味着人工智能决策越接近于合谋。下面，我们对智能体训练结果进行展示。

##### 1. $\beta=0.2$ 的情况

表-2 展示了在 $\beta=0.2$ 时智能体研发决策的联合训练结果。首先可以看到的是，智能体研发决策的结果普遍地偏离了理论模型中的纳什均衡，即形成了一定程度的“合谋”，这与现

有文献的研究结论相一致。下面，我们对训练结果进行具体介绍，并阐述初步的发现。

第一，随着智能体数量的增加，CI 值下降，NE 实现次数具有增加趋势，CE 实现次数显著下降。首先，从 CI 值上看：（1）就  $CI_{rd}$  来讲，在学习率  $\theta=0.25$  时，两智能体联合训练为 0.5249，三智能体联合训练为 0.5116，四智能体联合训练为 0.4386，具有明显的下降趋势，且这种趋势在  $\theta=0.50$  和  $\theta=0.75$  时也成立。（2）就  $CI_{pr}$  来讲，在学习率  $\theta=0.25$  时，两智能体联合训练为 0.7102，四智能体联合训练时则下降为 0.6712，尽管在三智能体联合训练时提高到 0.7145，但整体上呈现出下降趋势，且这种趋势在  $\theta=0.50$  和  $\theta=0.75$  也存在。其次，从均衡实现次数上看：（1）就 NE 实现次数来讲，在学习率为  $\theta=0.50$  时，二、三和四智能体联合训练分别实现了 12，7 和 42 次；在学习率为  $\theta=0.75$  时，二、三和四智能体联合训练分别实现了 9，18 和 89 次，上升趋势是很明显的。但在习率为  $\theta=0.25$  时，则未能发现此种趋势。（2）就 CE 实现次数来讲，在学习率  $\theta=0.25$  时，两智能体联合训练为 12 次，三智能体联合训练为 2 次，四智能体联合训练为 0 次，具有明显的下降趋势，且这种趋势在  $\theta=0.50$  和  $\theta=0.75$  时也成立。

表-2  $\beta=0.2$  时智能体研发决策联合训练结果

N	$\theta$	观测指标	均值	标准误	95%置信区间		CI	均衡实现次数	
					下界	上界		NE	CE
2	0.25	研发投入	28.57	0.5742	27.43	29.70	0.5249	12	12
		利润水平	373.48	3.8823	365.78	381.18	0.7102		
	0.50	研发投入	28.99	0.5703	27.86	30.12	0.5024	12	13
		利润水平	371.58	3.8594	363.92	379.23	0.6945		
	0.75	研发投入	28.33	0.5416	27.26	29.40	0.5374	9	13
		利润水平	376.74	3.5529	369.69	383.79	0.7372		
3	0.25	研发投入	20.49	0.4011	19.70	21.29	0.5116	2	2
		利润水平	176.67	2.9025	170.92	182.43	0.7586		
	0.50	研发投入	20.71	0.4770	19.76	21.65	0.4999	7	3
		利润水平	171.15	3.5047	164.19	178.10	0.7145		
	0.75	研发投入	21.71	0.4829	20.75	22.67	0.4449	18	1
		利润水平	163.90	4.2693	155.43	172.37	0.6567		
4	0.25	研发投入	17.05	0.3675	16.32	17.78	0.4386	10	0
		利润水平	92.15	2.9649	86.27	98.03	0.6712		
	0.50	研发投入	19.73	0.4574	18.83	20.64	0.2712	42	0
		利润水平	65.96	3.8667	58.29	73.64	0.4149		
	0.75	研发投入	23.24	0.2517	22.74	23.74	0.0525	89	0
		利润水平	33.89	2.2555	29.41	38.36	0.1010		

第二，随着学习率  $\theta$  的提高，CI 值下降，非合作均衡实现次数显著提高而合作均衡的实现次数显著下降。首先，从合谋指数上看：（1）就研发投入来讲，在智能体数量  $N=3$  和  $N=4$  时， $CI_{rd}$  随着  $\theta$  从 0.25 提高到 0.75 都出现了明显的下降趋势，仅在  $N=2$  时， $CI_{rd}$  出现了先下降再上升的波动。（2）就利润水平来讲， $CI_{pr}$  的变化具有同样的趋势。当智能体数量  $N=2$ ，在学习率  $\theta$  为 0.25，0.50 和 0.75 时， $CI_{pr}$  分别为 0.7102，0.6945 和 0.7372，即呈现出先下降后又提高的趋势。当智能体数量为 3 和 4 时，随着  $\theta$  的提高， $CI_{pr}$  则是不断下降的。其次，从均衡实现次数上看：（1）就 NE 实现次数来讲，除在  $N=2$  时， $\theta$  的提高并未引起非合作均衡实现次数的增加外，在  $N=3$  和  $N=4$  两种情况下，随  $\theta$  的提高，非合作均衡实现次数都有显著增加。（2）就 CE 实现次数来讲，未能发现其随  $\theta$  改变出现明显变化。

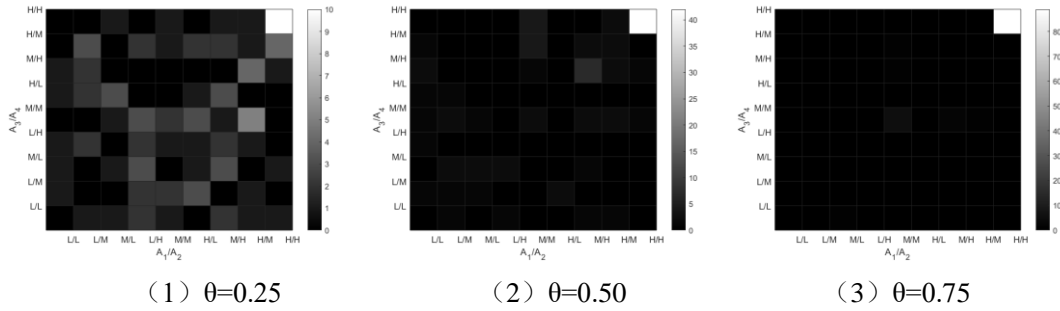


图-2  $\beta=0.2$  时四智能体联合训练不同行动组合实现频次

为了更直观地展示智能体训练结果，也更全面地披露在 100 次训练中各个行动组合的实现次数，我们以  $N=4$  为例，绘制不同行动组合实现频次图（图-2）。该图横轴表示智能体 1 和智能体 2 的行动组合（即  $A_1/A_2$ ），纵轴表示智能体 3 和智能体 4 的行动组合（即  $A_3/A_4$ ）。智能体有三种行动选择，即低研发投入（L）、中研发投入（M）和高研发投入（M）。因此，每两个智能体有 9 种行动组合，4 个智能体共有 81 种行动组合，图中每个单元格表示一种行动组合，左下角的单元格就是所有智能体都选择低研发投入的行动组合，也即“合谋”均衡，右上角的单元格表示所有智能体都选择高研发投入的行动组合，也即纳什均衡，其他单元格的含义依次类推。另外，该图以颜色的深浅表示行动组合的实现次数，颜色越暗，表示实现次数越少，反之则越多。从图中我们可以发现，随着  $\theta$  的提高，非合作均衡实现次数显著提高，但其他策略组合的实现次数并未展示出明显的规律性。在后面的分析中，我们还将 在  $\beta=0.3$  和  $\beta=0.4$  时展示该图，从而直观地检验  $\beta$  变化对智能体行动选择的影响。

### 2. $\beta=0.3$ 的情况

表-3 展示了在  $\beta=0.3$  时智能体研发决策的联合训练结果。体智能体研发决策的结果仍然普遍偏离纳什均衡结果，但与  $\beta=0.2$  的情况相比，训练结果收敛到 NE 的次数更少，训练结果随智能体数量和学习率变化的显著性有所下降。下面，我们进行具体介绍。

第一，智能体数量变化产生的影响。首先，从 CI 值上看：（1）就  $CI_{rd}$  来讲，在  $\theta=0.25$  时， $N=2$ ， $N=3$  和  $N=4$  三种情况下分别为 0.5025，0.5550 和 0.4875，呈现出先上升后下降的趋势。在  $\theta=0.50$  时，则是先下降再上升；在  $\theta=0.75$  时，又变为先上升后下降。（2）就  $CI_{pf}$  来讲，在  $\theta=0.25$  时，随智能体数量增加先上升后下降；在  $\theta=0.50$  时，随智能体数量增加而不断上升；在  $\theta=0.75$  时，随智能体数量增加先上升而后下降。其次，从均衡实现次数上看：

（1）就 NE 实现次数来讲，在  $\theta=0.25$  和  $\theta=0.50$  时，随智能体数量增加，NE 实现次数存在一定提升趋势，但在  $\theta=0.75$  时， $N=2$ ， $N=3$  和  $N=4$  时分别为 12，1 和 24，整体上呈现出波动上升的趋势。（2）就 CE 实现次数来讲，随智能体数量增加出现较为明显的下降趋势。在  $\theta=0.25$  时， $N=2$ ， $N=3$  和  $N=4$  时分别为 6，9 和 0 次；在  $\theta=0.50$  时， $N=2$ ， $N=3$  和  $N=4$  时分别为 12，0 和 1 次；在  $\theta=0.75$  时， $N=2$ ， $N=3$  和  $N=4$  时分别为 12，5 和 1 次。

表-3  $\beta=0.3$  时智能体研发决策联合训练结果

N	$\theta$	观测指标	均值	标准误	95%置信区间		CI	均衡实现次数	
					下界	上界		NE	CE
2	0.25	研发投入	30.09	0.3670	29.36	30.82	0.5025	11	6
		利润水平	425.44	1.7556	421.96	428.93	0.6825		
	0.50	研发投入	30.37	0.4255	29.53	31.21	0.4825	15	12
		利润水平	422.84	1.9668	418.93	426.74	0.6398		
	0.75	研发投入	29.99	0.4186	29.15	30.82	0.5100	12	12
		利润水平	424.45	1.8772	420.73	428.18	0.6662		
3	0.25	研发投入	21.00	0.3466	20.31	21.69	0.5550	3	9

		利润水平	211.72	1.5878	208.57	214.87	0.7343		
	0.50	研发投入	22.53	0.3031	21.93	23.14	0.4467	2	0
		利润水平	205.62	1.5531	202.53	208.70	0.6445		
	0.75	研发投入	21.38	0.2230	20.94	21.82	0.5283	1	5
		利润水平	211.09	1.4186	208.27	213.90	0.7250		
4	0.25	研发投入	17.23	0.24	16.76	17.70	0.4875	2	0
		利润水平	123.55	1.16	121.26	125.85	0.7055		
	0.50	研发投入	17.23	0.27	16.70	17.76	0.4875	2	1
		利润水平	122.63	1.27	120.10	125.15	0.6896		
	0.75	研发投入	18.27	0.37	17.53	19.00	0.4063	24	1
		利润水平	114.98	2.05	110.91	119.04	0.5576		

第二，学习率  $\theta$  提高产生的影响。首先，从 CI 值上看：（1）就  $CI_{rd}$  来讲，在  $N=2$  和  $N=3$  时，随学习率增加先下降后上升；在  $N=4$  时，随学习率增加而不断下降。（2）就  $CI_{pf}$  来讲，其变动趋势与研发投入时的变化趋势基本一致，即在  $N=2$  和  $N=3$  时，随学习率的提高先下降后上升；在  $N=4$  时，随学习率的提高而不断下降。其次，从均衡实现次数上看：（1）就 NE 实现次数来讲，在  $N=2$  和  $N=3$  时的变化随  $\theta$  变化的趋势并不明显；在  $N=4$  时，当  $\theta=0.25$ ， $\theta=0.50$  和  $\theta=0.75$  时，实现次数分别为 2，2 和 24，呈现出较为明显的增加趋势。（2）CE 实现次数并未随学习率的变化产生具较为明显的变化趋势。

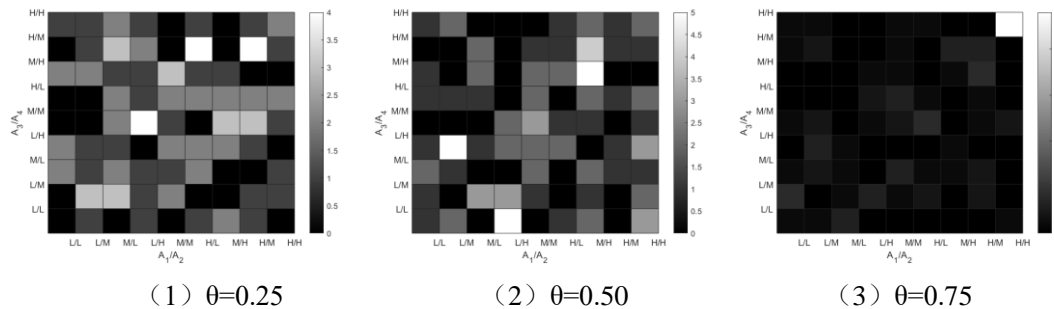


图-3  $\beta=0.3$  时四智能体联合训练不同行动组合实现频次

图-3 绘制了  $\beta=0.3$  时四智能体联合训练不同行动组合实现频次。从图中可以看出，当学习率  $\theta=0.75$  时，NE 出现的次数显著的增加。但在  $\theta=0.25$  和  $\theta=0.50$  两种情况中，并未发现具有明显更高实现次数的行动组合，说明智能体行动选择的收敛性受到较强随机因素的影响。与  $\beta=0.2$  时相比，81 种行动组合的出现频次更加均匀，仅在  $\theta=0.75$  时，NE 才具有明显的优势。

### 3. $\beta=0.4$ 的情况

表-4 展示了在  $\beta=0.4$  时智能体研发决策的联合训练结果。与  $\beta=0.2$  和  $\beta=0.3$  两种情况相比较， $\beta=0.4$  时的 CI 值整体上有所提升，CE 实现次数也有所增加，这表明  $\beta$  的提高增强了智能体的“合作”倾向。下面，我们进行具体分析。

第一，智能体数量变化产生的影响。首先，从 CI 值上看：（1）就  $CI_{rd}$  来讲，在  $\theta=0.25$  时， $N=2$ ， $N=3$  和  $N=4$  三种情况下  $CI_{rd}$  分别为 0.5225，0.5200 和 0.5225，存在先下降后上升的趋势，且这一趋势在在  $\theta=0.50$  和  $\theta=0.75$  两种情况下也存在。（2）就  $CI_{pf}$  来讲，在  $\theta=0.25$  时， $N=2$ ， $N=3$  和  $N=4$  三种情况下分别为 0.6484，0.6586 和 0.6619，存在上升趋势。但这种趋势在  $\theta=0.50$  和  $\theta=0.75$  时变为先下降后上升。其次，从均衡实现次数上看：（1）就 NE 实现次数来讲，在  $\theta=0.25$ ， $\theta=0.50$  和  $\theta=0.75$  三种情况下，随着智能体数量的增加，NE 实现次数均呈下降趋势，这与  $\beta=0.2$  和  $\beta=0.3$  时完全相反。（2）就 CE 实现次数来讲，在  $\theta=0.25$ ，

$\theta=0.50$  和  $\theta=0.75$  三种情况下，CE 的实现次数都是不断下降的。

表-4  $\beta=0.4$  时智能体研发决策联合训练结果

N	$\theta$	观测指标	均值	标准误	95%置信区间		CI	均衡实现次数	
					下界	上界		NE	CE
2	0.25	研发投入	31.30	0.2299	30.85	31.76	0.5225	9	14
		利润水平	486.03	0.5003	485.04	487.02	0.6484		
	0.50	研发投入	31.30	0.20	30.91	31.70	0.5225	6	9
		利润水平	486.27	0.4377	485.40	487.14	0.6624		
	0.75	研发投入	31.15	0.2315	30.69	31.61	0.5425	8	15
		利润水平	486.35	0.4945	485.36	487.33	0.6666		
3	0.25	研发投入	22.90	0.1964	22.51	23.29	0.5200	4	4
		利润水平	255.58	0.4906	254.60	256.55	0.6586		
	0.50	研发投入	23.06	0.2172	22.63	23.49	0.5217	5	5
		利润水平	255.16	0.5477	254.07	256.25	0.6383		
	0.75	研发投入	23.17	0.1973	22.78	23.56	0.4883	5	2
		利润水平	254.96	0.5018	253.96	255.95	0.6285		
4	0.25	研发投入	18.05	0.1608	17.73	18.37	0.5225	1	4
		利润水平	157.01	0.3729	156.27	157.75	0.6619		
	0.50	研发投入	17.96	0.1718	17.62	18.30	0.5338	1	4
		利润水平	157.23	0.3929	156.45	158.10	0.6738		
	0.75	研发投入	18.00	0.1418	17.72	18.28	0.5288	1	0
		利润水平	157.39	0.3274	156.74	158.04	0.6826		

第二，学习率  $\theta$  提高产生的影响。首先，从 CI 值上看：（1）就  $CI_{rd}$  来讲，在  $n=2$  时，随  $\theta$  的提高不断提高；在  $N=3$  和  $N=4$  时，随  $\theta$  的提高而先上升后下降。（2）就  $CI_{pf}$  来讲，在  $N=2$  和  $N=4$  时，随  $\theta$  的提高不断提高；但在  $N=3$  时，随  $\theta$  的提高而不断下降。其次，从均衡实现次数上看：（1）就 NE 实现次数来讲，在  $N=2$  时，实现次数随  $\theta$  提高先下降后上升；在  $N=3$  时，实现次数随  $\theta$  提高逐渐上升；在  $N=4$  时，实现次数保持 1 未变。（2）就 CE 实现次数来讲，在  $N=2$  时，实现次数随  $\theta$  提高先下降后上升；在  $N=3$  时，实现次数随  $\theta$  提高先上升后下降；在  $N=4$  时，实现次数随  $\theta$  提高先保持未变，后下降为 0。

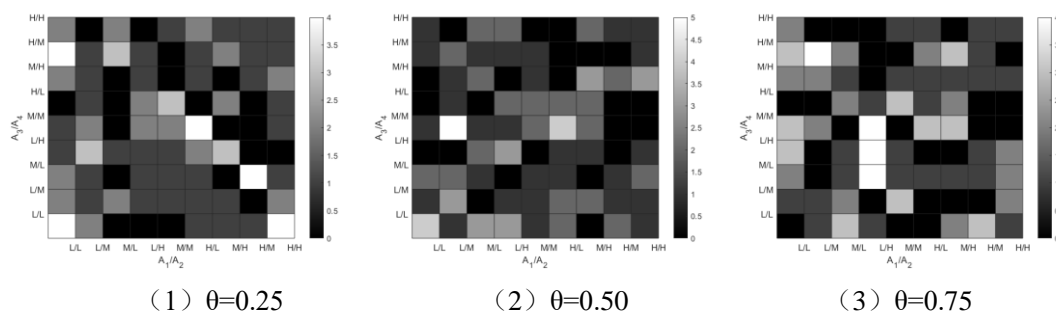


图-4  $\beta=0.4$  时四智能体联合训练不同行动组合实现频次

图-4 绘制了  $\beta=0.4$  时四智能体联合训练不同行动组合实现频次。从图中能够观察到 CE 的实现次数在  $\theta=0.75$  是出现了明显下降，其他行动组合的出现频率具有较强随机性。我们将图-2，图-3 和图-4 进行整体比较，能够发现随着  $\beta$  值的提高，NE 的实现次数有明显减少，而 CE 的实现次数有明显的增加，同时智能体行动组合收敛方向愈加的发散。下面，我们将对智能体训练结果进行统计学分析，以得到更加明确的研究结论。

## 五、对智能体决策影响因素的统计分析

### 1. 整体回归分析

为了在统计学意义上考察  $\beta$ ,  $N$  和  $\theta$  对智能体决策的影响, 我们将共计 2700 次的智能体训练结果作为统计样本, 设定计量模型如下:

$$DV_i = \alpha_0 + \alpha_1\beta_i + \alpha_2N_i + \alpha_3\theta_i + \varepsilon_i \quad (I)$$

其中,  $DV$  为被解释变量, 分为四种设定, 即研发投入合谋指数 ( $CI_{rd}$ ), 利润合谋指数 ( $CI_{pf}$ ), 是否实现了非合作均衡 (INE) 以及是否实现了合作均衡 (ICE)。其中, INE 和 ICE 为 0-1 变量, 因此我们采用 logit 模型进行回归分析。解释变量即为研发外溢程度 ( $\beta$ ), 智能体数量 ( $N$ ) 和学习率 ( $\theta$ )<sup>⑤</sup>。 $\varepsilon$  为随机扰动项。

表-5  $\beta$ ,  $N$  和  $\theta$  对智能体决策的影响

	(1) $CI_{rd}$	(2) $CI_{pf}$	(3) INE	(4) ICE
$\beta$	0.0050*** (0.0006)	0.0026*** (0.0007)	0.9026*** (0.0088)	1.0148 (0.0109)
$N$	-0.0510*** (0.0064)	-0.0528*** (0.0072)	1.6106*** (0.1476)	0.2809*** (0.0403)
$\theta$	-0.0012*** (0.0002)	-0.0018*** (0.0003)	1.0283*** (0.0035)	0.9991 (0.0043)
Cons	0.5407*** (0.0299)	0.8166*** (0.0320)	0.1116*** (0.0423)	1.1256 (0.6121)
样本数	2700	2700	2700	2700
R 平方	0.0548	0.0412	0.1344	0.0975

注: ①括号里为估计系数的异方差稳健标准误; ②\*, \*\*, \*\*\*分别表示回归系数在 10%, 5%和 1%水平上显著; ③第 (3) 和 (4) 列各单元格中第一行数字报告的是几率比。

表-5 报告了计量模型(I)的估计结果。第 (1) 列展示了  $CI_{rd}$  作为被解释变量的估计结果, 可以看出研发外溢度  $\beta$  对其具有显著的正向影响,  $\beta$  提高 0.01 导致  $CI_{rd}$  提高 0.005; 智能体数量  $N$  对  $CI_{rd}$  具有显著的负向影响, 智能体数量每增加 1 个将导致  $CI_{rd}$  下降 0.051; 学习率  $\theta$  对  $CI_{rd}$  产生显著的负向影响,  $\theta$  提高 0.01 将导致  $CI_{rd}$  下降 0.0012。第 (2) 列展示了  $CI_{pf}$  作为被解释变量的估计结果, 可以看出  $\beta$  对  $CI_{pf}$  具有显著的正向影响,  $\beta$  提高 0.01 导致  $CI_{pf}$  提高 0.0026; 智能体数量  $N$  对  $CI_{pf}$  具有显著的负向影响, 智能体数量每增加 1 个将导致  $CI_{pf}$  下降 0.0528; 学习率  $\theta$  对  $CI_{pf}$  产生显著的负向影响,  $\theta$  提高 0.01 将导致  $CI_{pf}$  下降 0.0018。第 (3) 列展示了被解释变量为 INE 的估计结果, 其中  $\beta$  的影响显著为负, 它每提高 0.01 将导致非合作均衡实现的概率下降 9.74%;  $N$  的影响显著为正, 每增加 1 个智能体将导致非合作均衡实现的概率提高 61.06%;  $\theta$  的影响显著为正, 它每提高 0.01 将导致非合作均衡实现的概率提高 2.83%。第 (4) 列展示了被解释变量为 ICE 的估计结果, 其中  $\beta$  的影响为正向但不显著; 智能体数量的影响显著为负, 每增加 1 个智能体将导致非合作均衡实现的概率下降 71.91%; 学习率的影响为负向, 但并不显著。

### 2. 依参数取值分组后智能体决策结果与各参数的相关性分析

尽管在上面的整体回归分析中我们得到了各参数对智能体决策的影响, 但第四部分的结果展示告诉我们, 这种影响很可能不是全局性的。因此, 为了更加精确地考察各参数对智能体决策的影响, 我们在不同的参数取值组合下对各参数与智能体决策结果进行逐一的相关性分析, 这能够让我们看到各参数发挥作用的区间, 从而帮助我们更全面地考察各参数对

<sup>⑤</sup> 为了是估计结果更易理解, 我们将  $\beta$  和  $\theta$  乘以 100 后的数值作为解释变量 Beta 和 Theta。

智能体决策结果的协同影响。

(1) 智能体决策结果与  $\beta$  的相关性。表-6 展示了在  $N$  和  $\theta$  的不同组合下, 智能体决策结果与  $\beta$  的相关性分析。首先,  $CI_{rd}$  除在  $N=2$  且  $\theta=0.25$  时与  $\beta$  负相关外, 在其他的  $N$  和  $\theta$  组合下均与  $\beta$  正相关, 但仅在  $N=4$  时, 相关关系显著。第二,  $CI_{pf}$  与  $\beta$  在  $N=2$  和  $N=3$  时负相关, 且在  $N=2$  且  $\theta=0.25$ ,  $N=3$  且  $\theta=0.25$  以及  $N=3$  且  $\theta=0.50$  三种情况下负相关关系显著。但这种负相关关系在  $N=4$  且  $\theta=0.50$  和  $N=4$  且  $\theta=0.75$  两种情况下转为正相关, 在  $N=4$  且  $\theta=0.75$  时正相关关系显著。第三,  $INE$  与  $\beta$  在  $N=3$  且  $\theta=0.25$  时正相关, 在其他情况下均为负相关, 在  $N=3$  且  $\theta=0.75$  和  $N=4$  是负相关关系显著。第四,  $ICE$  与  $\beta$  在  $N=2$  且  $\theta=0.50$  时为负相关, 在  $N=4$  且  $\theta=0.75$  无相关关系, 在其他情况下均为正相关, 在  $N=4$  且  $\theta=0.25$  和  $N=4$  且  $\theta=0.50$  两种情况下相关关系显著。

表-6 智能体决策结果与  $\beta$  的相关性

	$CI_{rd}$ 与 $\beta$	$CI_{pf}$ 与 $\beta$	$INE$ 与 $\beta$	$ICE$ 与 $\beta$	样本数
$n=2, \theta=0.25$	-0.0034	-0.0844	-0.0397	0.0265	100
$n=2, \theta=0.50$	0.0286	-0.0439	-0.0783	-0.0515	100
$n=2, \theta=0.75$	0.0071	-0.0974*	-0.0138	0.0240	100
$n=3, \theta=0.25$	0.0147	-0.1722***	0.0479	0.0375	100
$n=3, \theta=0.50$	0.0029	-0.1198**	-0.0387	0.0507	100
$n=3, \theta=0.75$	0.0728	-0.0425	-0.1956***	0.0253	100
$n=4, \theta=0.25$	0.1631***	-0.0161	-0.1805***	0.1424**	100
$n=4, \theta=0.50$	0.4031***	0.3442	-0.4688***	0.1276**	100
$n=4, \theta=0.75$	0.6570***	0.6575***	-0.7401***	0.0000	100

注: \*, \*\*, \*\*\*分别表示回归系数在 10%, 5%和 1%水平上显著(下同)。

(2) 智能体决策结果与  $n$  的相关性。表-7 展示了在  $\beta$  和  $\theta$  的不同组合下, 智能体决策结果与  $N$  的相关性分析。首先,  $CI_{rd}$  与  $N$  在  $\beta=0.3$  且  $\theta=0.50$ ,  $\beta=0.4$  且  $\theta=0.25$  和  $\beta=0.4$  且  $\theta=0.50$  等三种情况下正相关, 在其他情况下负相关, 在  $\beta=0.2$  且  $\theta=0.25$ ,  $\beta=0.2$  且  $\theta=0.50$ ,  $\beta=0.2$  且  $\theta=0.75$  和  $\beta=0.3$  且  $\theta=0.75$  等四种情况下负相关关系显著。第二,  $CI_{pf}$  与  $N$  在  $\beta=0.3$  且  $\theta=0.25$ ,  $\beta=0.3$  且  $\theta=0.50$ , 以及  $\beta=0.4$  时正相关, 在其他情况下负相关, 在  $\beta=0.2$  且  $\theta=0.50$ ,  $\beta=0.2$  且  $\theta=0.75$  和  $\beta=0.3$  且  $\theta=0.75$  时负相关关系显著。第三,  $INE$  与  $N$  在  $\beta=0.2$  且  $\theta=0.50$ ,  $\beta=0.2$  且  $\theta=0.75$ ,  $\beta=0.3$  且  $\theta=0.75$  等三种情况下具有显著的正相关关系, 在其他情况下负相关, 在  $\beta=0.3$  且  $\theta=0.25$ ,  $\beta=0.3$  且  $\theta=0.50$  以及  $\beta=0.4$  时负相关关系显著。第四,  $ICE$  在九种情况下均与  $N$  成负相关关系, 但在  $\beta=0.4$  且  $\theta=0.50$  时不显著。现有文献对市场竞争强度如何影响企业研发投入主要有三种差异性观点<sup>[52]</sup>, 一是垄断势力更强的企业更有激励进行技术创新<sup>[53,54]</sup>; 二是激烈的市场竞争迫使企业加大研发投入从而提高生产效率<sup>[55-57]</sup>; 三是市场竞争强度与企业研发之间存在非线性关系<sup>[58,59]</sup>。对算法决策而言, 随着智能体数量的增加, 研发合谋的倾向会下降, 企业的研发投入会增加, 因此, 算法决策更支持第二种理论观点。

表-7 智能体决策结果与  $n$  的相关性

	$CI_{rd}$ 与 $n$	$CI_{pf}$ 与 $n$	$INE$ 与 $n$	$ICE$ 与 $n$	样本数
$\beta=0.2, \theta=0.25$	-0.1377**	-0.0562	-0.0301	-0.2323***	100
$\beta=0.2, \theta=0.50$	-0.3122***	-0.3224***	0.3043***	-0.2362***	100
$\beta=0.2, \theta=0.75$	-0.6177***	-0.6436***	0.6707***	-0.2516***	100
$\beta=0.3, \theta=0.25$	-0.0262	0.0389	-0.1635***	-0.1124*	100
$\beta=0.3, \theta=0.50$	0.0083	0.0781	-0.2179***	-0.2206***	100

$\beta=0.3, \theta=0.75$	-0.1532***	-0.1466**	0.1499***	-0.1891***	100
$\beta=0.4, \theta=0.25$	0.0000	0.0224	-0.1548***	-0.1566***	100
$\beta=0.4, \theta=0.50$	0.0187	0.0192	-0.1042*	-0.0860	100
$\beta=0.4, \theta=0.75$	-0.0231	0.0273	-0.1355**	-0.2649***	100

(3) 智能体决策结果与  $\theta$  的相关性。表-8 展示了在  $\beta$  和  $N$  的不同组合下, 智能体决策结果与  $\theta$  的相关性分析。首先,  $CI_{rd}$  与  $\theta$  在  $N=2$  和  $\beta=0.4$  且  $N=4$  时为正相关但不显著, 在其他情况下为负相关, 在  $\beta=0.2$  且  $N=3$ ,  $\beta=0.2$  且  $N=4$  和  $\beta=0.3$  且  $N=4$  等三种情况下负相关关系显著。第二,  $CI_{pf}$  与  $\theta$  在  $\beta=0.2$  且  $N=2$ ,  $\beta=0.4$  且  $N=2$  和  $\beta=0.4$  且  $N=4$  等三种情况下为正相关但不显著, 在其他情况下为负相关, 在  $\beta=0.2$  且  $N=3$ ,  $\beta=0.2$  且  $N=4$  和  $\beta=0.3$  且  $N=4$  等三种情况下负相关关系显著。第三, NE 实现次数与  $\theta$  在  $\beta=0.2$  且  $N=3$ ,  $\beta=0.2$  且  $N=4$  和  $\beta=0.3$  且  $N=4$  等三种情况下显著正相关, 在  $\beta=0.3$  且  $N=2$ ,  $\beta=0.4$  且  $N=3$  和  $\beta=0.4$  且  $N=4$  三种情况下正相关但不显著, 在其他情况下负相关但不显著。第四, CE 实现次数与  $\theta$  在  $\beta=0.4$  且  $N=4$  时显著负相关, 在其他情况下的相关性均不显著。

表-8 智能体决策结果与  $\theta$  的相关性

	$CI_{rd}$ 与 $\theta$	$CI_{pf}$ 与 $\theta$	INE 与 $\theta$	ICE 与 $\theta$	样本数
$\beta=0.2, n=2$	0.0172	0.0354	-0.0391	0.0123	100
$\beta=0.2, n=3$	-0.1088*	-0.1440**	0.2282***	-0.0292	100
$\beta=0.2, n=4$	-0.5669***	-0.6103***	0.6462***	0.0000	100
$\beta=0.3, n=2$	0.0106	-0.0217	0.0123	0.0816	100
$\beta=0.3, n=3$	-0.0470	-0.0167	-0.0583	-0.0774	100
$\beta=0.3, n=4$	-0.1419**	-0.2210***	0.3087***	0.0502	100
$\beta=0.4, n=2$	0.0291	0.0270	-0.0153	0.0123	100
$\beta=0.4, n=3$	-0.0529	-0.0495	0.0194	-0.0434	100
$\beta=0.4, n=4$	0.0125	0.0424	0.0000	-0.1014*	100

## 六、总结与讨论

### 1. 结论与贡献

依据第四部分的智能体训练结果和第五部分对训练结果进行的统计分析, 我们可以得到如下结论: 第一, 尽管算法决策结果向“合谋”收敛的概率并不高, 但与纳什均衡结果相比, 研发投入更低而利润水平更高的情况是普遍出现的。第二, 智能体数量的增加能够降低“合谋”实现的概率, 且这种影响具有较高的全局性。第三, 研发的外溢程度会促进智能体“合谋”, 但这种促进作用是局部的, 在智能体数量较多时更易实现。第四, 学习率的提高能够降低“合谋”实现的概率, 但这种作用也仅在局部发生, 在智能体数量较多且研发外溢性较低时作用更显著。

本文可能的贡献有以下几点: 第一, 由于企业研发行为存在外溢性, 本文分析了外溢程度的变化对智能体决策产生的影响, 这是对现有理论研究的一个重要补充。第二, 本文深入挖掘了影响算法合谋的三个重要因素, 以及各因素影响作用发挥的协同性, 一定程度上揭示了算法“合谋”的内在逻辑性。第三, 本文研究了运用机器学习算法进行研发投入决策的可能结果, 阐明人工智能的无限制应用可能会对企业创新产生负向影响, 证明了高科技成果的应用可能阻碍科技进步这一悖论的存在, 这对数字经济时代如何进一步推动企业加大研发力度, 提升研发效率具有较强的启示作用, 对算法合谋的规制政策制定也具有较大的借鉴意义。



## 2. 政策建议

依据本文研究结论,我们提出政策建议如下:第一,机器学习算法决策在竞争强度不同的行业的应用应差异化推进。由于参与博弈的智能体数量能够显著地降低“合谋”发生的概率,因此,在竞争性较强的行业,可放宽算法决策的准入门槛。而对于规模经济性较强,企业数量较少,竞争性较弱的行业,对算法决策应更加审慎,从而有效避免具有极大隐蔽性而易于逃避反垄断法监管的智能体“合谋”的发生。第二,进一步加强知识产权保护,降低由于知识产权保护不足导致的过度外溢性。知识产权保护不足所导致的过度知识外溢会使企业减少研发投入,这种影响机制在算法决策中同样存在。本文的研究表明,在算法研发决策中,过度的知识外溢不仅引起研发投入下降,还促进了智能体“合谋”从而进一步降低研发投入。因此,应进一步加强知识产权保护,避免知识过度外溢,减少研发竞争中算法“合谋”的发生。第三,对应用于企业决策的人工智能算法加强反垄断监管。本文研究结论发现,在智能体训练中的参数设置(如学习率 $\theta$ )也会影响算法“合谋”的发生概率。因此,应加强对应用于企业决策的人工智能算法的反垄断监管,对智能体的学习机制、参数设定等在不同的应用场景下进行系统研究,制定科学的算法监管办法。

## 3. 研究展望

人工智能在企业决策领域的应用具有巨大潜力,也为学术研究提供了广阔的空间。但由于本文研究边界和计算机算力等主客观条件的限制,我们将以下两个重要议题留待后续研究做进一步深入,也期待与学界同仁共同探讨:第一,本文没有考虑行业存在企业进入和退出的情况,也即参与研发竞争的智能体数量是既定不变的。然而,依据可竞争市场理论,在不存在沉没成本且企业能够自由地进入和退出一个行业时,即使一个行业仅有少数几个在位企业,博弈均衡也将是竞争性的。那么,如果在智能体训练环境的设定上加入潜在进入企业的威胁,是否能够对在位智能体的决策产生影响呢,这是需要进一步研究的重要问题。第二,对政府产业政策如何在算法决策中发挥影响的研究也有待展开。政府产业政策的影响可以是多方面的,如对消费领域的补贴可能导致市场整体需求的提升,对企业研发投入的补贴会降低企业成本,对企业“合谋”的处罚也必然引起智能体决策奖励信号的改变。后续研究可综合考察不同政策的实施效果,并对不同政策实施条件进行分析和比较。

## 参考文献:

- [1] Duso, T., Röller, L., Seldeslachts, J.. Collusion through Joint R&D: An Empirical Assessment. *The Review of Economics and Statistics*, 2014,96(2):349-370.
- [2] Sovinsky, M. Do Research Joint Ventures Serve A Collusive Function?. *Journal of the European Economic Association*, 2022,20(1):430-475.
- [3] 毕金平.论企业共同研发行为的反垄断豁免制度. *江淮论坛*, 2014,264(2):124-129.
- [4] 蒋逸舟,吕伟文.知识产权反垄断案件中“研发市场”的适用问题研究. *海南金融*, 2019, 369 (08): 45-50.
- [5] 孟昌,曲寒瑛. 算法合谋及其规制研究进展. *经济学动态*, 2021(6):128-143.
- [6] 苏敏,夏杰长. 数字经济中竞争性垄断算法合谋的治理困境. *财经问题研究*, 2021(11):37-46.
- [7] 乔榛,刘瑞峰. 大数据算法的价格合谋问题研究[J]. *学术交流*, 2022,334(01):82-95+191.
- [8] Schwalbe, U.. Algorithms, Machine Learning, and Collusion. *Journal of Competition Law & Economics*, 2019,14(4):568-607.
- [9] Gautier, A., Ittoo, A., Cleynenbreugel, P.V.. AI Algorithms, Price Discrimination and Collusion: A Technological, Economic and Legal Perspective. *European Journal of Law and Economics*, 2020,50:405-435.

- [10] Ezrachi,A., Stucke, M.E.. Sustainable and Unchallenged Algorithmic Tacit Collusion. *Northwestern Journal of Technology &Intellectual Property*, 2020,17(2):217-260.
- [11] Turner,D.. The Definition of Agreement under the Sherman Act: Conscious Parallelism and Refusals to Deal. *Harvard Law Review*, 1962,75(4):655-706.
- [12] 沈亮亮. 算法在市场竞争者中的应用与法律难题——从大数据杀熟谈起. *太原学院学报(社会科学版)*, 2019(3) : 26-31.
- [13] Posner,R.A.. Oligopoly and the Antitrust Laws: A Suggested Approach. *Stanford Law Review*, 1969,21(6):1562-1606.
- [14] Kaplow,L.. *Competition Policy and Price Fixing*. Princeton University Press, 2013.
- [15] 李振利,李毅. 论算法共谋的反垄断规制路径. *学术交流*, 2018(7):73-82.
- [16] 施春风. 定价算法在网络交易中的反垄断法律规制. *河北法学*, 2018(11):111-119.
- [17] Veljanovski,C.. Pricing Algorithms as Collusive Devices. <https://ssrn.com/abstract=3644360>, 2020.
- [18] Calvano,E., Calzolari, G., Denicolo, V., Pastorello, S.. Algorithmic Pricing What Implications for Competition Policy. *Review of Industrial Organization*, 2019, 55(1):155-171.
- [19] 唐要家,尹钰锋. 算法合谋的反垄断规制及工具创新研究. *产经评论*, 2020(3):5-16.
- [20] 尹振涛,陈媛先,徐建军.平台经济的典型特征、垄断分析与反垄断监管. *南开管理评论*, 2022,25(03):213-226.
- [21] 刘学. 人工智能时代算法合谋反垄断的适法困境与逻辑依归. *西北民族大学学报*, 2022(1):87-94.
- [22] Milgrom, Paul R., Tadelis, Steven. How Artificial Intelligence and Machine Learning Can Impact Market Design. NBER Working Paper, 2017, No.24282.
- [23] Assad, S., Clark, R., Ershov, D., Xu, L.. Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market. <https://papers.ssrn.com/abstract=3682021>, 2020.
- [24] Salcedo,B.. Pricing Algorithms and Tacit Collusion. Manuscript, PSU-Cornell 2015 Fall Macroeconomics Conference, 2016.
- [25] Zhou, N., Zhang, L., Li, S., Wang, Z.. Algorithmic Collusion in Cournot Duopoly Market: Evidence from Experimental Economics. <http://export.arxiv.org/pdf/1802.08061>, 2018.
- [26] Calvano, E., Calzolari, G., Denicolò, V., Pastorell,S.. Artificial Intelligence, Algorithmic Pricing, and Collusion. *American Economic Review*, 2020,110(10):3267-97.
- [27] Klein, T. Autonomous Algorithmic Collusion: Q-learning under Sequential Pricing. *Rand Journal of Economics*, 2021,52(3):538-599.
- [28] Meylahn, J M., Janssen, L.. Limiting Dynamics for Q-learning with Memory one in Symmetric Two-Player, Two-Action Games. *Complexity*, 2022, Article ID: 4830491. <https://doi.org/10.1155/2022/4830491>.
- [29] Beneke,F., Mackenrodt,M.O.. Artificial Intelligence and Collusion. *IIC-International Review of Intellectual Property and Competition Law*, 2019, 50:109-134.
- [30] Stucke,M.E., Ezrach, A.. How Pricing Bots could Form Cartels and Make Things More Expensive. <https://hbr.org/2016/10/how-pricing-bots-could-form-cartels-and-make-things-more-expensive>, 2016.
- [31] OECD.. Algorithms and Collusion: Competition Policy in the Digital Age[EB/OL]. <https://www.oecd.org/competition/algorithms-collusion-competition-policy-in-the-digital-age.htm>, 2017.

- [32] Ezrachi, A., Stucke, M.E.. Two Artificial Neural Networks Meet in An Online Hub and Change the Future (of Competition, Market Dynamics and Society). Oxford Legal Studies Research Paper, No. 2017/24.
- [33] Levenstein, M.C., Suslow, V.Y.. Breaking up is Hard to do: Determinants of Cartel Duration. *Journal of Law and Economics*, 2011,54(2):455-492.
- [34] Mehra, S.K.. Antitrust and the Robo-seller: Competition in the Time of Algorithms. *Minnesota Law Review*, 2016,204, <https://scholarship.law.umn.edu/mlr/204>.
- [35] Ballard,D.I., Naik, A.S.. Algorithms, Artificial Intelligence, and Joint Conduct. *CPI Antitrust Chronicle*, 2017,1(2), <https://www.competitionpolicyinternational.com/algorithms-artificial-intelligence-and-joint-conduct/>.
- [36] Deng,A.. What do We Know about Algorithmic Tacit Collusion. *Antitrust*, 2018,33(1):88-95.
- [37] Gal, M.S.. Algorithms as Illegal Agreements. *Berkeley Technology Law Journal*, 2019,34(1):67-118.
- [38] Harrington, J.E.. Developing Competition Law for Collusion by Autonomous Artificial Agents. *Journal of Competition Law & Economics*, 2019,14(3):331-363.
- [39] Gal, M. S., Elkin-Korren, N.. Algorithmic Consumers. *Harvard Journal of Law & Technology*, 2017,30(2):309-353.
- [40] Airoidi, M., Rokka, J.. Algorithmic Consumer Culture, *Consumption Markets & Culture*, 2022,25(5): 411-428.
- [41] Marcellis-Warin, N.D., Marty, F., Thelisson, E. Warin., T... Artificial intelligence and Consumer Manipulations: from Consumer's Counter Algorithms to Firm's Self-regulation Tools. *AI and Ethics*,2022, 2: 259–268.
- [42] Watkins, C.J.C.H., Dayan, P.. Q-learning. *Machine Learning*, 1992, 8: 279-292.
- [43] Waltman, L., Kaymak, U.. Q-Learning Agents in a Cournot Oligopoly Model. *Journal of Economic Dynamics and Control*, 2008,32 (10): 3275-93.
- [44] Ittoo,A., Petit,N.. Algorithmic Pricing Agents and Tacit Collusion: A Technological Perspective. In: J. Herve & D.S. Alexandre (eds). *L' intelligence artificielle et le droit*, Bruxelles: Larcier, 2017.
- [45] CMA.. Pricing Algorithms. <https://www.covcompetition.com/2018/11/the-cmas-paper-on-pricing-algorithms-collusion-and-personalised-pricing/>,2018.
- [46] Mnih, V., Kavukcuoglu, K., Silver, D., Andrei, A., Rusu, J. V., Bellemare, M.G., Graves, A.. Human-Level Control through Deep Reinforcement Learning. *Nature*, 2015,518 (7540): 529-33.
- [47] Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L, Driessche, G., Schrittwieser, J.. Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature*, 2016,529 (7587): 484-89.
- [48] Silver, D., Hubert,T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M.. A General Reinforcement Learning Algorithm that Masters Chess, Shogi, and Go through Self-Play. *Science*, 2018,362 (6419): 1140-44.
- [49] Watkins, C. J. C. H.. Learning from Delayed Rewards. PhD diss. King's College, 1989.
- [50] Kaelbling, L.P., Littman, M.L., Moore, A.W.. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 1996,4: 237-285.
- [51] Sandholm, T.W., Crites, R.H.. Multiagent Reinforcement Learning in the Iterated Prisoner's Dilemma. *Biosystems*, 1996,37: 147-166.

- [52] 郭晓玲, 李凯. 不同竞争模式下制造商创新投入决策影响研究. 东北大学学报 (自然科学版), 2022,43 (1) : 124-132.
- [53] Aghion, P., Howitt, P.. A Model of Growth through Creative Destruction. *Econometrica*, 1992,60(2):323-351.
- [54] Blundell,R., Griffith,R., Reenen, J.V.. Market Share, Market Value and Innovation in A Panel of British Manufacturing Firms. *Review of Economic Studies*, 1999, 66(3):529-554.
- [55] Arrow, K.J.. *Economic Welfare and the Allocation of Resources for Inventions, the Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton: Princeton University Press, 1962:609-626.
- [56] Scherer, F.M.. Market Structure and the Employment of Scientists and Engineers. *American Economic Review*, 1967,57(3):524-531.
- [57] Gilbert, R.J., Newbery, D.M.G.. Preemptive Patenting and the Persistence of Monopoly. *American Economic Review*, 1982,74(3):514-526.
- [58] Aghion, P., Bloom, N., Griffith, R., Howitt, P., Blundell, R.W.. Competition and Innovation: An Inverted-u Relationship. *Quarterly Journal of Economics*, 2005, 120(2):701-728.
- [59] Thshler, A., Milstein, I.. R&D Wars and the Effects of Innovation on the Success and Survivability of Firms in Oligopoly Markets. *International Journal of Industrial Organization*, 2009,27(4):519-531.

## Algorithms Decision-making and Corporate R&D "Collusion"

Xu Lei

(School of Economics, Liaoning University, Shenyang, 110036)

**Abstract:** This paper sets up an agent decision-making environment for enterprise R&D game, uses reinforcement learning Q-learning algorithm to train agents, and analyzes and predicts the possible results caused by artificial intelligence empowering enterprise R&D. The results show that: first, although the probability of convergence of the agent's decision-making results to "collusion" is not high, it is common for lower R&D investment and higher profit levels than Nash equilibrium results. Second, the increase in the number of agents can reduce the probability of achieving "collusion", and this effect has a high global character. Third, the spillover degree of R&D will promote the "collusion" of agents, but this promotion effect is local and easier to achieve when the number of agents is large. Fourth, the increase in the learning rate can reduce the probability of achieving "collusion", but this effect only occurs locally, and it is more obvious when the number of agents is large and the spillover of R&D is low. Based on the above research results, this paper puts forward three suggestions to strengthen the management of artificial intelligence application at the decision-making level of enterprises, which provides theoretical support for the continuous improvement of anti-monopoly regulation in the era of digital economy.

**Key Words:** Artificial Intelligence; Q-learning Algorithm; R&D Game; Enterprise Innovation